DOCUMENT RESUME

ED 206 672 TM 810 567

AUTHOR Kalsbeek, William D.; And Others
TITLE No-Show Analysis. Final Report.
INSTITUTION Research Triangle Inst., Durham, N.C. Statistics Research Div.
SPONS AGENCY Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.; National Center for Education Statistics (DHEW), Washington, D.C.
REPORT NO NAEP-2550-1061-3
PUB DATE Apr 75
CONTRACT OEC-0-74-0506
NOTE 46p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Assessment; High Schools; Mathematical Models; *Performance Factors; Sampling; *Statistical Analysis
IDENTIFIERS *National Assessment of Educational Progress; *Nonresponders; Second Science Assessment (1973)

ABSTRACT

The National Assessment of Educational Progress: Second Science Assessment No-Show Study assessed the magnitude and causation of nonresponse biases. A No-Show is defined as an individual who was selected as a sample respondent but failed to be present for regular assessment of the 17-year-old group. The procedure whereby a sample of eligible 17-year-old No-Shows were selected to take four specific No-Show assessment packages is briefly described. Estimates of biases due to nonresponse were made for the following domain variables: region; sex; race; size and type of community: derived parental education. These domains are outlined. Also documented are the domain estimation methodologies, terminology and the methods utilized for the computation of formulas. No-Show domain analysis results are briefly summarized in tables, but are detailed more fully elsewhere in separate appendixes. Results show that reliability estimates were positive and several were significant, indicating that regular assessment students generally performed better on assessment packages than did the no-shows. Separate exercise analyses of bias are considered in section 6 as an extension of the analyses covered by the study. Primary type of information provided by report: Procedures (Evaluation) (Sampling). (Author/AEF)

255U-1061-3

FINAL REPORT

NO-SHOW ANALYSIS

by

William D. Kalsbeek
Anne F. Clemmer
Ralph E. Folsom, Jr.

Statistics Research Division
Research Triangle Institute

Prepared for

National Assessment of Educational Progress

April 1975

2

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

"The work upon which this publication is based was performed pursuant to a Contract with the Education Commission of the States, utilizing funds from the U.S. Office of Education, Department of Health, Education and Welfare, Contract No. OEC-0-74-0506. However, the opinions expressed herein do not necessarily reflect the position or policy of the U.S. Office of Education or the Education Commission of the States, and no official endorsement by the U.S. Office of Education or the Education Commission of the States should be inferred."

1. BACKGROUND AND OBJECTIVES

The analysis of National Assessment No-Shows (i.e., nonrespondents to package administrations) was undertaken to assess the magnitude of associated nonresponse biases as well as to determine the origins and causation of these biases. Field operations were conducted both In-School and Out-of-School. A total of 1990 packages were administered during the In-School portion of the study. Auxiliary data on students' attendence records, course grades and curricula were collected on 1753 Age Class 3 regular assessment paticipants and 1324 No-Shows. Packages were also administered Out-of-School to 130 of the 598 No-Shows who could not be located during an In-School followup. Finally, supplementary data, regarding the reasons for student absenteeism from regular assessment administrations, were ascertained from 1989 In-School No-Shows and 130 Out-of-School No-Shows.

For the purposes of this study a No-Show is defined as an individual who was initially selected as a sample respondent but failed to be present for seventeen-year-old regular assessment. The No-Show study was conducted for seventeen-year-old respondents only since it was felt that the nonresponse problem was most serious at that age level. A multistage probability sample was selected from the NAEP Year 04 In-School sample. Briefly, the sample was selected as follows. First, a sample of primary units was selected from the NAEP Year 04 In-School primary sample. Secondly, within each selected No-Show PSU, those schools eligible to take a No-Show package were determined. Thirdly, within these eligible schools, eligible 17-year-old No-Shows were selected to take specific No-Show packages. Finally, a systematic sample of these In-School

No-Shows who could not be contacted in school was selected for an Out-of-School followup study. Further details of sampling and field procedures are presented elsewhere [1]. Procedures for computing sample weights from the No-Show study sample design have also been presented [2].

Four packages were selected for the No-Show study. The four packages included three group packages, selected from the eleven Year 04 group packages; and one individual package, selected from the three Year 04 individual packages. The selected group packages were numbers 1, 3, and 9; and the selected individual package was number 13.

Auxiliary and supplementary data results have been documented [3]. Among several other possible observations are the following:

(1) No-Shows, as compared to regular assessment students, are absent from school more often, get lower course grades, are less inclined to enroll in college preparatory curricula, and enroll in fewer courses;

(2) About half of In-School No-Shows claimed that they were not notified about the regular assessment administration;

(3) About half (of those In-School No-Shows who remembered) were absent from school on their original regular assessment administration data;

(4) The morning hours of the day were clearly preferred for regular assessment sessions;

(5) About half of the Out-of-School No-Shows were not enrolled in the school where NAEP had assumed that they were enrolled;

(6) A moderate proportion of those In-School No-Shows who said they were not absent from school on the day of regular assessment indicated that they had other school-oriented commitments.

A discussion of methodology and a presentation of accompanying estimates of package composite biases and rel-biases have been documented [4-5]. Most estimates were positive and several were significant, indicating that regular assessment students generally performed better on these packages than did the No-Shows. The above-mentioned auxiliary data results tend to parallel these findings. With the group packages, the magnitude of bias was smaller and statistically significance less often when only In-School No-Shows were involved in the computations.

The overall purpose of the present document will be to present further methodology and results. Specifically, estimates of biases due to nonresponse are made for several domains or subpopulations. The domains considered here are formulated from several variables which are of interest to NAEP and which have been determined for each sampled regular assessment and No-Show student. The variables considered here are as follows:

(1) Region;

(2) Sex;

(3) Race;

(4) Size and Type of Community (STOC);

(5) Derived Parental Education.

A discussion of the specific domains, formed as marginal categories of the above variables, is presented in section 2 of this report.

The measurement variable for a package in these analyses is the proportion of mathematics or science exercises in the package that are answered correctly by a regular assessment or No-Show respondent. The total number of exercises associated with the mathematics and science portions of the three group packages (01, 03, and 09), that are considered in these analyses, is presented in table 3 of appendix B. This measurement variable yields a simple average of the separate-exercise P-values involved. Similar separate-exercise analyses of bias will be considered as an extension of the present analyses (see section 6.2).

It should also be mentioned initially that only the No-Show study group packages are considered in these analyses. In other words, No-Show study individual package 13 will not be included in these analyses. It is excluded because of the complex nature of the administration of the package. More specifically, it contains a large amount of conditional branching among certain exercises making the identification of "key" exercises for evaluation a difficult task. On the other hand, exercises of the group packages are straight-forward and "key" exercises can be easily identified for evaluation purposes.

2. DESCRIPTION OF DOMAINS

Results for the National Assessment of Educational Progress surveys are published separately for each Age Class--9-year-olds, 13-year-olds, 17-year-olds, and young adults aged 26 to 35. Within each of these Age Classes, results are reported for each of the five groups or domains. The domain variables are sex, region of country, race, Size and Type of Community (STOC), and derived parental education. The domain variables along with the reporting categories for each domain variable are listed in table 2.1. The tables in appendix A compare 17-year-old regular assessment respondents with 17-year-old nonrespondents or No-Shows by domain. The nonrespondents could have been contacted either through the In-School or Out-of-School portion of the No-Show study. The No-Show sample and data collection activities have been documented elsewhere [1].

2.1 Sex Domains

The two sexes, male and female, are the reporting categories for the sex domain. From appendix table A.1, it can be seen, in general, that there are an approximately equal number of male and female respondents and nonrespondents.

2.2 Region Domains

The four regional reporting categories are those regional groupings defined by the Office of Business Economics, Department of Commerce. The four regional groupings by State are defined in table 2.2. From appendix table A.3, it can be seen that the regular respondents are fairly equally divided between the four regions. Nonrespondents, on the other hand, are slightly more concentrated in the Central and Western regions. This event

Table 2.1. Domain variables and reporting categories for National Assessment of Educational Progress

Domain variable	Reporting categories
Sex	Male Female
Region	Northeast Southeast Central West
Race	White Black Other
Size and Type of Community (STOC)	Extreme rural Low metropolitan High metropolitan Inner city fringe Suburban fringe Medium city Small city
Derived parental education	No high school Some high school Graduated from high school Post high school

PAGE 7 TABLE 2.2

OMMITTED DUE TO CONFIDENTIALITY

occurred because of the refusal of several schools in the Northeast and Southeast to release the names and addresses of nonrespondents who had not been contacted during the In-School portion of the No-Show study.

2.3 Race Domains

Race reporting categories are white, black, and other. The "other" category includes Puerto Ricans, Mexicans, Orientals, Polynesians, Asians, American Indians, etc. Observing appendix table A.2, it can be seen that approximately 70 percent of the sample respondents and nonrespondents are white, approximately 15 percent are black, and approximately 10 percent are classified as other.

2.4 STOC Domains

The seven STOC categories are a means of classifying respondents by characteristics of their home community such as size, location, and occupation of residents. The first two categories, extreme rural and low metropolitan, are oversampled to obtain enough respondents to make domain estimates with the desired precision. A sufficient number of respondents in the third category, high metropolitan, are obtained by normal sampling procedures to make domain estimates with the desired precision. A detailed explanation of the definition, formation, and size of STOC categories is provided elsewhere [6]. Observing appendix table A.4, one sees that approximately 40 percent of the sample respondents and nonrespondents are classified in the small city STOC category; approximately 10 percent are classified in each of the remaining six STOC categories.

2.5 Derived Parental Education Domains

Derived parental education codes were determined by comparing the highest grade of school completed for both parents and selecting the highest grade among both parents. If this highest grade completed was at most 8, then the derived parental education classification was no high school; if this highest grade completed was at least 9 but less than 12, then the classification was some high school; if this highest grade completed was 12, then the classification was graduated from high school; and finally, if this highest grade completed was greater than 12, then the classification was post high school. Referring to appendix table A.5, it can be noted that approximately 10 percent of the sample respondents and nonrespondents did not answer this quesiton either because they did not know the highest education level of both parents, they refused or forgot to answer the question, or because of a specific State law prohibiting a research organization from asking a student this question. Approximately 10 percent of the sample respondents and nonrespondents had derived parental education classifications of no high school; approximately 10 percent were classified as some high school; about 30 percent were classified as graduated from high school; and about 40 percent were classified as post high school.

3. DESCRIPTION OF METHODOLOGY

3.1 Definitions and Terminology

The notation used in this paper is similar to that used in a previous paper which reported bias estimates by subject and package as found in the No-Show study. This paper was presented at the August 1974 meeting of the American Statistical Association [5].

The formulas in this paper are developed conditional upon the selection of Primary Sampling Units (PSU's); they are also developed specific to a particular domain category, and specific to a particular subject matter area within the package. A symbol defines an entity, while the attached subscript serves to determine its applicability. A block symbol refers to a random variable, and a script symbol refers to a parameter. Finally, an upper case script symbol refers to the population of all units and the corresponding lower case script symbol refers to an estimate of the parameter associated with a sample of these units. Specifically, we define

Y = proportion of exercises answered correctly,

$\mathcal{P}(p)$ = population (sample) proportion of eligible NAEP participants,

$\mathcal{E}(e)$ = population (sample) number of eligible students.

The first-position subscript (α) to be associated with the above symbols will refer to the total population (o), regular assessment respondents (1), or nonrespondents or No-Shows (2). Let $X_{\alpha jk}$ be a domain indicator variable such that

$$X_{\alpha jk} = \begin{cases} 1, & \text{if eligible student-}k\text{ belonging to response group-}\alpha \\ & \text{in school-}j\text{ is a member of the specified domain;} \\ 0, & \text{otherwise.} \end{cases}$$

We then define

$$YX_{\alpha j} = \sum_{k=1}^{E_{\alpha j}} X_{\alpha jk} Y_{\alpha jk} / E_{\alpha j}$$

and

$$X_{\alpha j} = \sum_{k=1}^{E_{\alpha j}} X_{\alpha jk} / E_{\alpha j} .$$

Population totals $F_\alpha(yx)$, $C_\alpha(yx)$, $F_\alpha(x)$, and $C_\alpha(x)$ will refer to the quantities

$$F_\alpha(yx) = \sum_{j\epsilon\Omega} E_{oj} P_{1j} YX_{\alpha j};\ \alpha = 1,2$$

$$C_\alpha(yx) = \sum_{j\epsilon\Omega} E_{oj} P_{2j} YX_{\alpha j};\ \alpha = 1,2$$

$$F_\alpha(x) = \sum_{j\epsilon\Omega} E_{oj} P_{1j} X_{\alpha j};\ \alpha = 1,2$$

$$C_\alpha(x) = \sum_{j\epsilon\Omega} E_{oj} P_{2j} X_{\alpha j};\ \alpha = 1,2 .$$

Sample estimates for these quantities will be denoted by $f_\alpha(yx)$, $c_\alpha(yx)$, $f_\alpha(x)$, and $c_\alpha(x)$ respectively. The quantities will be combined to assess the magnitude of nonresponse bias by domain in NAEP regular assessment statistics.

The following symbols are used in the preceding and subsequent [illegible] formulation:

h = pseudo-stratum,

i = PSU within pseudo-stratum,

j = school,

k = student within school,

m. = number of eligible sample students taking a package,

w = package sample nonresponse adjusted weight (i.e., inverse of the probability of selection into the study),

Ω = set of all eligible schools,

ω = sample set of eligible schools,

+ = summation over all possible subscript values.

3.2 Domain Estimation Methodology

3.2.1 First-Order PSU Estimators

First, note that the "true" domain mean $\overline{YX}_o$ is

$$\overline{YX}_o = \frac{\sum_{j\in\Omega} E_{oj} [P_{1j}\, YX_{1j} + P_{2j}\, YX_{2j}]}{\sum_{j\in\Omega} E_{oj} [P_{1j}\, X_{1j} + P_{2j}\, X_{2j}]}$$

$$= \frac{P_1(yx) + C_2(yx)}{P_1(x) + C_2(x)}.$$

The regular assessment estimator for $\overline{YX}_o$ is

$$\hat{\overline{yx}}_1 = \frac{\sum_{j\in\omega_1} \sum_{k=1}^{m_{1j}} w_{1jk}\, X_{1jk}\, Y_{1jk}}{\sum_{j\in\omega_1} \sum_{k=1}^{m_{1j}} w_{1jk}\, X_{1jk}}$$

with expectation

$$E(\overline{yx}_1) = \frac{\sum_{j\varepsilon\Omega} E_{oj}\cdot YX_{1j}}{\sum_{j\varepsilon\Omega} E_{oj} X_{1j}}$$

$$= \frac{F_1(yx) + C\ (yx)}{F_1(x) + C_1(x)}$$

Therefore,

$$\text{Bias } (\overline{yx}_1) = E(\overline{yx}_1) - \overline{YX}_o \tag{3.2.1.1}$$

$$= \frac{F_1(yx) + C_1(yx)}{F_1(x) + C_1(x)} - \frac{F_1(yx) + C_2(yx)}{F_1(x) + C_2(x)}$$

Similarly,

$$\text{Rel-Bias } (yx_1) = \text{Bias } (\overline{yx}_1) / \overline{YX}_o \tag{3.2.1.2}$$

$$= \left\{ \left[\frac{F_1(yx) + C_1(yx)}{F_1(yx) + C_2(yx)}\right]\left[\frac{F_1(x) + C_2(x)}{F_1(x) + C_1(x)}\right] - 1 \right\}$$

Ratio-type estimators are used to estimate values associated with equations 3.2.1.1 and 3.2.1.2. If we let

$$\rho = f_1(yx) + c_1(yx), \tag{3.2.1.3}$$

$$\sigma = f_1(x) = c_1(x), \tag{3.2.1.4}$$

$$\tau = f_1(yx) + c_2(yx), \tag{3.2.1.5}$$

$$\upsilon = f_1(x) + c_2(x), \tag{3.2.1.6}$$

then

$$\text{bias } (\overline{yx})_1 = \frac{\rho}{\sigma} - \frac{\tau}{\upsilon}$$

and

$$\text{rel-bias } (\overline{yx}_1) = \frac{\rho\upsilon}{\sigma\tau} - 1 ,$$

where

$$f_1(yx) = \sum_{j\epsilon\omega_1} P_{1j} \sum_{k=1}^{m_{1j}} w_{1jk} X_{1jk} Y_{1jk} ,$$

$$f_1(x) = \sum_{j\epsilon\omega_1} P_{1j} \sum_{k=1}^{m_{1j}} w_{1jk} X_{1jk} ,$$

$$c_1(y) = \sum_{j\epsilon\omega_1} P_{2j} \sum_{k=1}^{m_{1j}} w_{1jk} X_{1jk} Y_{1jk} ,$$

$$c_1(x) = \sum_{j\epsilon\omega_1} P_{2j} \sum_{k=1}^{m_{1j}} w_{1jk} X_{1jk} ,$$

$$c_2(yx) = \sum_{j\epsilon\omega_2} \sum_{k=1}^{m_{2j}} w_{2jk} X_{2jk} Y_{2jk} ,$$

$$c_2(x) = \sum_{j\epsilon\omega_2} \sum_{k=1}^{m_{2j}} w_{2jk} X_{2jk} .$$

The parameters P_{1j} and P_{2j} are estimated from school response rates during regular assessment. The estimate for a No-Show study group package is found as the response rate to all group packages given in that school. The w_{2jk}

weights denote the reciprocals of No-Show selection probabilities adjusted for No-Show nonresponse.

The preceding statistics yield domain bias estimates involving In-School regular assessment respondents and all No-Show respondents. Another set of meaningful domain bias estimates involves In-School regular assessment respondents and In-School No-Show respondents. The definition changes indicated by the (*) in equations 3.2.1.7 through 3.2.1.10 were motivated by an attempt to form a matched school domain bias estimator based exclusively on In-School No-Shows. The set of schools (ω_1^*) is the subset of regualr assessment schools (ω_2) which provided In-School No-Show responses for the particular package in question. The deleted schools either had no cooperating In-School No-Show respondents for the package, or were subsampled out at the No-Show package assignment stage to control the package yield per PSU. The regular assessment respondents for the set of ω_1^* schools with In-School No-Show responses for the package were inflated to account for the deleted schools, hence the adjusted w_{1jk}^* weights. Thus,

$$\text{bias}^*(\overline{yx}_1) = \frac{\rho^*}{\delta^*} - \frac{\tau^*}{\upsilon^*}$$

and

$$\text{rel-bias}^*(\overline{yx}_1) = \frac{\rho^*\upsilon^*}{\delta^*\tau^*} - 1 ,$$

where

$$\rho^* = f_1^*(yx) + c_1^*(yx) \quad ; \tag{3.2.1.7}$$

$$\delta^* = f_1^*(x) + c_1^*(x) \quad , \tag{3.2.1.8}$$

$$\tau^* = f_1^*(yx) + c_2^*(yx) \quad , \tag{3.2.1.9}$$

$$\upsilon^* = f_1^*(x) + c_2^*(x) \quad , \tag{3.2.1.10}$$

and

$$f_1^*(yx) = \sum_{j\varepsilon\omega_1^*} P_{1j} \sum_{k=1}^{m_{1j}} w_{1jk}^* X_{1jk} Y_{1jk} \quad ,$$

$$f_1^*(x) = \sum_{j\varepsilon\omega_1^*} P_{1j} \sum_{k=1}^{m_{1j}} w_{1jk}^* X_{1jk} \quad ,$$

$$c_1^*(yx) = \sum_{j\varepsilon\omega_1^*} P_{2j} \sum_{k=1}^{m_{1j}} w_{1jk}^* X_{1jk} Y_{1jk} \quad ,$$

$$c_1^*(x) = \sum_{j\varepsilon\omega_1^*} P_{2j} \sum_{k=1}^{m_{1j}} w_{1jk}^* X_{1jk} \quad ,$$

$$c_2^*(yx) = \sum_{j\varepsilon\omega_1^*} \frac{P_{2j}\, m_{1j}}{m_{2j}^*} \sum_{k=1}^{m_{2j}^*} w_{1jk}^* X_{2jk}^* Y_{2jk}^* \quad ,$$

$$c_2^*(x) = \sum_{j\varepsilon\omega_1^*} \frac{P_{2j}\, m_{1j}}{m_{2j}^*} \sum_{k=1}^{m_{2j}^*} w_{1jk}^* X_{2jk}^* \quad .$$

With m^*_{2j} denoting the number of In-School No-Show responses from school $j\epsilon\omega^*_1$, the definition of the set of schools ω^*_1 assures that $m^*_{2j} > 0$.

3.2.2 Overall First-Order Estimates

To facilitate the ensuing discussion, attach subscripts to ρ, δ, τ, and υ of formula 3.2.1.3 through 3.2.1.6 and ρ^*, δ^*, τ^*, and υ^* of formula 3.2.1.7 through 3.2.1.10 (i.e., subscripts "hi" to indicate PSU-i within pseudo-stratum-h). Using these quantities, one obtains the overall estimate involving all No-Shows as

$$\text{bias}(\overline{yx}_1) = \frac{\rho_{++}}{\delta_{++}} - \frac{\tau_{++}}{\upsilon_{++}} \tag{3.2.2.1}$$

and

$$\text{rel-bias}(\overline{yx}_1) = \frac{\rho_{++}\upsilon_{++}}{\delta_{++}\tau_{++}} - 1 \tag{3.2.2.2}$$

and, involving only In-School No-Shows, as

$$\text{bias}^*(\overline{yx}_1) = \frac{\rho^*_{++}}{\delta^*_{++}} - \frac{\tau^*_{++}}{\upsilon^*_{++}} \tag{3.2.2.3}$$

and

$$\text{rel-bias}^*(\overline{yx}_1) = \frac{\rho_{++}\upsilon_{++}}{\delta^*_{++}\tau^*_{++}} - 1 \tag{3.2.2.4}$$

[illegible]

[illegible]

3.2.3 Second-Order Estimators

The second-order estimators of variance for expressions 3.2.2.1 through 3.2.2.4 are based upon a form of the "jackknife" technique introduced by Quenouille [7] and advanced for interval estimation by Tukey [8]. The procedure is presented for domain estimates involving all No-Shows, although the procedure for domain estimates involving only In-School No-Shows is similar.

First, we define

$$\beta_{h1} = 2\left[\frac{\rho_{++}}{\delta_{++}} - \frac{\tau_{++}}{\upsilon_{++}}\right] - \left[\frac{\rho_{++} + \rho_{h1} - \rho_{h2}}{\delta_{++} + \delta_{h1} - \delta_{h2}} - \frac{\tau_{++} + \tau_{h1} - \tau_{h2}}{\upsilon_{++} + \upsilon_{h1} - \upsilon_{h2}}\right],$$

$$\beta_{h2} = 2\left[\frac{\rho_{++}}{\delta_{++}} - \frac{\tau_{++}}{\upsilon_{++}}\right] - \left[\frac{\rho_{++} + \rho_{h2} - \rho_{h1}}{\delta_{++} + \delta_{h2} - \delta_{h1}} - \frac{\tau_{++} + \tau_{h2} - \tau_{h1}}{\upsilon_{++} + \upsilon_{h2} - \upsilon_{h1}}\right],$$

$$\gamma_{h1} = 2\left[\frac{\rho_{++}\,\upsilon_{++}}{\delta_{++}\,\tau_{++}} - 1\right] - \left[\frac{(\rho_{++} + \rho_{h1} - \rho_{h2})(\upsilon_{++} + \upsilon_{h1} - \upsilon_{h2})}{(\delta_{++} + \delta_{h1} - \delta_{h2})(\tau_{++} + \tau_{h1} - \tau_{h2})} - 1\right],$$

$$\gamma_{h2} = 2\left[\frac{\rho_{++}\,\upsilon_{++}}{\delta_{++}\,\tau_{++}} - 1\right] - \left[\frac{(\rho_{++} + \rho_{h2} - \rho_{h1})(\upsilon_{++} + \upsilon_{h2} - \upsilon_{h1})}{(\delta_{++} + \delta_{h2} - \delta_{h1})(\tau_{++} + \tau_{h2} - \tau_{h1})} - 1\right].$$

Since the 57 PSU's make up a half-sample of NAEP regular assessment PSU's, the desirable condition of having two PSU selections per stratum does not hold. Instead, pseudo strata were formed by sequentially pairing the No-Show PSU's according to region and size. Since the number of PSU's is odd, one pseudo stratum (h^o) was assigned three PSU's. The associated jackknife estimators of variance are

$$\text{ver}\{\text{bias}(\overline{yx}_1)\} = 1/4 \sum_{\substack{h=1 \\ h \neq h^o}}^{H} \left[\beta_{h1} - \beta_{h1}\right]^2 + 1/8 \sum_{i=1}^{2} \sum_{j=i+1}^{3} \left[\beta_{h^o i} - \beta_{h^o j}\right]^2$$

and

$$\text{var}\{\text{rel-bias}(\overline{yx}_1)\} = 1/4 \sum_{\substack{h=1 \\ h \neq h^o}}^{H} \left[\gamma_{h1} - \gamma_{h2}\right]^2 + 1/8 \sum_{i=1}^{2} \sum_{j=i+1}^{3} \left[\gamma_{h^o i} - \gamma_{h^o j}\right]^2 \ldots$$

To assess the significance of the domain bias and domain rel-bias estimates, one might be willing to assume that

$$T = \frac{\text{bias}(yx_1)}{\left[\text{var}\{\text{bias } \overline{yx}_1\}\right]^{1/2}}$$

and

$$T' = \frac{\text{rel-bias}(\overline{yx}_1)}{\left[\text{var}\{\text{rel-bias}(\overline{yx}_1)\}\right]^{1/2}}$$

for each domain category are distributed as "Student's" t-statistic with 29 degrees of freedom. Under this assumption, significance with a Type I error of 0.05 is indicated when $|T| \geq 2.045$ or $|T'| \geq 2.045$.

4. DESCRIPTION OF METHODOLOGY APPLICATION

The procedures for computing the formula described in section 3 are summarized in the four paragraphs which follow. A listing of the software developed to execute these computations is included as appendix B. The calculations described in the following paragraphs are performed for each No-Show group package (numbers 1, 3, and 9) and for each subject matter area (mathematics and science) within each group package. Before any production runs were made, all calculations were verified by hand.

The computation process followed several steps. First, those components of ρ, σ, τ, υ, ρ^*, σ^*, τ^*, and υ^* described in section 3.2.1, which pertain to regular assessment respondents and No-Show respondents, were computed separately for each PSU and for each domain category. Second, the quantities described in the preceding paragraph were combined to form ρ, σ, τ, υ, ρ^*, σ^*, τ^*, and υ^* for each PSU and for each domain category. These quantities were summed over all PSU's for each domain to generate ρ_{++}, σ_{++}, τ_{++}, υ_{++}, ρ^*_{++}, σ^*_{++}, τ^*_{++}, and υ^*_{++} described in section 3.2.2. Third, bias and rel-bias estimates were calculated for each domain category for all No-Shows and for In-School No-Shows only. These quantities have been previously described in section 3.2.2. In addition, β_{h1}, β_{h2}, γ_{h1}, and γ_{h2} of section 3.2.3 were computed for each pseudo-stratum-hi, for each domain category, and for all No-Shows and for In-School No-Shows only. Finally, the estimated variance of the bias and the rel-bias as described in section 3.2.3, was obtained for all No-Shows and for In-School No-Shows only for each domain category.

5. NO-SHOW DOMAIN ANALYSIS CONCLUSIONS

5.1 Tables Summarizing Results

The results of No-Show domain analyses are summarized in appendices C and D. Average per exercise estimates of bias and rel-bias by domain category are provided by package and subject matter area for all No-Shows and In-School No-Shows only in tables C.1 through C.6. The variance of these bias and rel-bias measures is also provided in the tables. Using the t-test described in section 3.2.3, those values of the bias and rel-bias which are significant with the probability of a Type I error of 0.05 were denoted by an asterisk.

Weighted estimates of the mean number correct responses per exercise by domain are provided in tables C.7 through C.12. These estimates are provided for regular assessment respondents as well as for In-School and Out-of-School No-Show respondents. Results are further delineated by package and subject matter area. The formula used to compute these estimates was

$$\bar{y}_\alpha = \frac{\sum_{j\epsilon\omega_\alpha} \sum_{k=1}^{m_{\alpha j}} w_{\alpha jk} X_{\alpha jk} Y_{\alpha jk}}{\sum_{j\epsilon\omega_\alpha} \sum_{k=1}^{m_{\alpha j}} w_{\alpha jk} X_{\alpha jk}}$$

Tables C.13 through C.15 provide weighted estimates of response rates by domain for regular assessment respondents. These estimates are provided separately by package. The formula used to compute these estimates was

$$\overline{P}_1 = \frac{\sum\limits_{j\varepsilon\omega_1} P_{1j} \sum\limits_{k=1}^{m_{1j}} w_{1jk} X_{1jk}}{\sum\limits_{j\varepsilon\omega_1} \sum\limits_{k=1}^{m_{1j}} w_{1jk} X_{1jk}}$$

Tables D.1 and D.2 provide estimates of the mean number correct responses by package and subject matter area for Age Class 3 Dropouts and Out-of-School No-Show respondents, respectively.

Table D.3 summarizes by domain the number of times the rel-bias for all No-Shows was significant and the number of times it was negative by subject matter area. The rel-bias measures are also ranked by size over all three No-Show group packages.

Lastly, table D.4 presents the average Out-of-School percentage contribution to the total bias. The average Out-of-School percentage contribution to the total bias was obtained as follows. First, the Out-of-School component of the bias was computed for each of the three No-Show group packages by subject matter area as

$$\text{Bias}(\overline{Y}_1) = (1-\overline{P}_1)\ [\underbrace{W(\overline{Y}_1-\overline{Y}_{2I})}_{\substack{\text{In-School}\\ \text{component}}} + \underbrace{(1-W)\ (\overline{Y}_1-\overline{Y}_{2\emptyset})}_{\substack{\text{Out-of-School}\\ \text{component}}}]$$

where

w = the average proportion of In-School No-Shows;

$\overline{Y}_1$ = the average number of correct responses for regular assessment respondents;

$\bar{Y}_{2I}$ = the average number of correct responses for In-School No-Shows;

$\bar{Y}_{2\emptyset}$ = the average number of correct responses for Out-of-School No-Shows.

Second, the Out-of-School percentage contribution was computed for each package as

$$\text{Out-of-School percentage} = \frac{\text{Out-of-School component}}{\text{In-School component + Out-of-School component}} \times 100\%.$$

Finally, the average Out-of-School percentage contribution was obtained by averaging the preceding quantity over all three group packages in the No-Show study.

5.2 Discussion of Results

As expected, the number of significant biases among domain estimates of the various bias measures in tables C.1 through C.6, is greater when all No-Shows are used than when In-School No-Shows only are used. The preceding statement is true for both bias and rel-bias measures. There are slightly more significant values with mathematics than science exercises within packages. Package 1 exercises tend to yield slightly more significant values. Since the performance determining exercises are assigned to packages arbitrarily, it is not surprising that the level of bias would vary from package to package.

Few consistent significance patterns emerge within package and among domains. With respect to rel-bias estimates, when all No-Shows are used, only Race—White exhibits significance with all combinations of subject matter and packages. Conversely, Race—Other and DPE—Some-High-School exhibit no significances. Results are no more consistent with corresponding bias measures.

The magnitude of the per exercise bias measure indicate that, in general, among domains, bias for mathematics exercises is slightly greater than for science exercises. Inconsistencies of this pattern are often attributable to small sample sizes.

Several patterns in magnitude emerge within domain variable categories. These patterns tend to be more pronounced with science exercises. For example, with rel-bias and all No-Shows in science, Sex--Male biases tend to be greater than Sex--Female biases. Region--West biases tend to be largest and Region--Northeast tends to be smallest. Race--Black is large while Race--White and Race--Other are smaller. STOC--Suburban-Fringe tends to be largest and STOC--Innter-City-Fringe smallest for science exercises. DPE--No-High-School tends to be the greatest while DPE--Post-High-School tends to be smallest.

The STOC patterns for mathematics exercises are more intuitive with Inner City Fringe and Small City yielding the largest biases and Suburban Fringe the smallest. Low Metro has relatively large biases for both science and mathematics exercises. Except for STOC, where the No-Show sample sizes for the extreme categories are very small, the above-mentioned patterns remain essentially intact for corresponding mathematics domains. A summary of some results from tables C.1 through C.6 is found in table D.3

Negative estimates tend to appear periodically but without much consistency with respect to domains. Negative estimates of bias measures, using only In-School No-Shows, tend to appear most frequently in STOC categories and are generally nonsignificant and small in size. Since sample sizes are often relatively small under these circumstances, one might attribute the negativeness to sampling error.

When one observes the results in tables C.7 through C.15 juxtaposed with those in tables C.1 through C.6, it becomes readily evident that the magnitude of domain biases in the later tables is directly related to the domain-specific difference in regular assessment and No-Show performances and indirectly related to domain-specific response rates. Once again, small sample sizes may disrupt the consistency of such a statement.

By observing tables D.1 and D.2 one notes that the performances of Age Class 3 Dropouts and Out-of-School No-Shows are virtually the same for all combinations of exercise subject matter and packages. This has administrative implications which are discussed subsequently.

Out-of-School No-Shows present a major contribution to the total bias. Results of table D.4 show that, in virtually all domains, Out-of-School No-Shows make up a small proportion (1-W) of all No-Shows, and yet they are responsible for a large share of the total bias. Since the results of tables D.1 and D.2 indicate that Out-of-School No-Shows and supplementary frame Dropouts perform similarly on the three group packages, NAEP might well adopt a nonresponse imputation procedure in which Dropouts assume the performance of Out-of-School No-Shows, or vice versa. Administratively, this implies that either Dropouts or Out-of-School No-Shows would be followed up and tested in order to reduce nonresponse bias. Assessment performance by imputation could be done in any of several ways. Subsequent analyses will be designed to investigate the effects of these imputation procedures on the magnitude of nonresponse bias. These analyses are discussed in section 6. The results discussed above would tend to support other findings in which the magnitude and number of significant results associated with biases are reduced when only In-School No-Shows are used in the estimate.

The size of the percentage of the Out-of-School contribution (see table D.4) in those domains with significant bias, does not seem to conform to any particular pattern when the bias is measured and all No-Shows are used. However, when rel-bias is considered and only In-School No-Shows are used, the percentages associated with significant domains tend to be relatively smaller, particularly in science domains. These phenomena would conform to intuition since significance with only In-School No-Shows would require a relatively large contribution by In-School No-Shows overall (i.e., low Out-of-School percentages).

Differences among average Out-of-School percentages within domain variables appear to be more heavily attributable to differences in W (proportion In-School No-Shows) than to absolute performance differences. This is observed in table D.4 by noticing a high negative correlation between the estimated average package value of W and the average Out-of-School percentage. The negative relationship can be explained from the formula used to calculate the Out-of-School percentage. Further details pertaining to this formula have been presented in section 5.1. For fixed $(\overline{Y}_1 - \overline{Y}_{2I})$ and $(\overline{Y}_1 - \overline{Y}_{2\emptyset})$, as W increases the Out-of-School contribution percentage decreases. For example, where the average Out-of-School percentage for STOC—Low-Metro is high, the corresponding value of W is relatively low.

The average of Out-of-School bias percentages differs markedly only with corresponding STOC categories between mathematics and science. This may be due to small sample sizes in STOC categories (see table A.4). With other domain variables differences between corresponding mathematics and science domains are small.

5.3 Recommendations

Two basic recommendations eminate from the findings of this study to date. First, since a large portion of No-Show bias is attributable to the Out-of-School No-Shows and since the performance of these No-Shows appears to be similar to Dropout frame respondents, NAEP may wish to investigate the possible use of Dropouts to compensate for bias of estimates that is attributable to Dropouts or Out-of-School No-Shows. This investigation could be in connection with an investigation of imputation procedures which are discussed in section 6.4.

Second, to reduce the number of In-School No-Shows, it is recommended that several alternative In-School followup procedures be investigated and compared. This recommendation arises from the fact that supplementary data have indicated that a large proportion of No-Shows had legitimate scheduling difficulties because of other school-related activities or illness. These findings are reinforced by the "Reason for No-Show" verification performed by the District Supervisor after nonresponse. Those results are presented in table 5.1. Perhaps greater flexibility in scheduling regular assessment administrators would realize a greater response among In-School No-Shows. To implement this recommendation, a small pilot study will be conducted in connection with Year 06 data collection to test results from a number of alternative followup procedures. This study is more fully described in section 6.1.

It is strongly recommended that every effort be made to maximize the proportion of initially selected students that receive their assigned regular

Table 5.1 "Reason for No-Show" codes
(verified by D.S. shortly after nonrepsonse)

Reason	Region 1		Region 2		Region 3		Region 4		Total	
Late for session	31	(3.6%)	20	(2.5%)	9	(1.1%)	22	(2.6%)	82	(2.5%)
Extra curricular activity	14	(1.6%)	23	(2.9%)	16	(1.9%)	17	(2.1%)	70	(2.1%)
Exam or important class	10	(1.2%)	9	(1.2%)	12	(1.4%)	25	(3.0%)	56	(1.7%)
Notified but forgot	25	(2.9%)	5	(0.6%)	6	(0.7%)	7	(0.9%)	43	(1.3%)
Went home sick	0	-	5	(0.6%)	7	(0.8%)	4	(0.5%)	16	(0.5%)
Late for school	10	(1.1%)	2	(0.3%)	6	(0.7%)	3	(0.4%)	21	(0.6%)
Work study program	19	(2.2%)	33	(4.2%)	27	(3.2%)	9	(1.1%)	88	(2.7%)
Report for work	10	(1.1%)	5	(0.6%)	9	(1.1%)	13	(1.6%)	37	(1.1%)
Cannot remember	3	(0.4%)	0	-	9	(1.1%)	26	(3.1%)	38	(1.1%)
Other	178	(20.5%)	64	(8.2%)	64	(7.6%)	173	(20.9%)	479	(14.4%)
Absent from school entire day	173	(19.9%)	290	(36.9%)	201	(23.7%)	250	(30.2%)	914	(27.5%)
Unknown	346	(39.8%)	237	(30.2%)	365	(43.1%)	173	(20.9%)	1,121	(33.7%)
Withdrawal from school	49	(5.7%)	90	(11.5%)	111	(13.1%)	102	(12.3%)	352	(10.5%)
Blank	0	-	2	(0.3%)	4	(0.5%)	3	(0.4%)	9	(0.3%)
TOTAL	868	(100.0%)	785	(100.0%)	846	(100.0%)	827	(100.0%)	3,326	(100.0%)

Table 5.1 (Continued)

Reason	Region 1	2	3	4	Total
Vacation	0	0	0	1	1
No Show	20	0	0	0	20
Discharged	7	0	0	0	7
Non-English	6	0	0	0	6
Truant	3	0	0	0	3
Refusal	69	5	11	10	95
Transferred	4	5	8	1	18
Dropout	11	15	13	18	57
Assessed	2	2	0	3	7
Homebound	1	4	4	3	12
Ineligible	1	3	2	11	17
Suspended	0	6	0	1	7
Graduated	0	4	2	2	8
Unknown/Blank	54	18	24	119	215
Not enrolled	0	1	0	0	1
Trade school	0	1	0	0	1
Check-out	0	0	0	4	4
TOTAL	178	64	64	173	479

assessment administrations. This recommendation implies that special efforts should be made by DS's to insure that school personnel notify each selected student of the date, time, and place of the administration. Exercise Administrators should allow for greater flexibility in scheduling administrations to allow for possible conflicts with other school-related activities. This procedure would include waiting a few extra minutes for late initially selected respondents to show up.

6. EXTENSIONS

Several other analyses could be undertaken in connection with the overall goals of the No-Show study. The purpose of this section is to preview some of these analyses.

6.1 Separate Exercise Analyses

Analyses of bias up until this point have been formulated at the level of a student's overall performance on the group of mathematics or science exercises found in each of the No-Show packages. As presently conceived, the large numbers of exercises which make up the packages would require that the scope of these analyses be limited to the total populations, and that if domain analyses were done, they would involve only a few "key" exercises.

Methodologically, the procedures and algorithms, which are used for three separate exercise analyses, will be similar to those used for previous analyses if the bias of the mean is considered. The difference is that the measurement variable is dichotomous under these circumstances, since, for a single exercise, the proportion of correct responses (Y, to follow the terminology of the present and past No-Show working papers) would be either zero or one. By this approach the bias of the mean proportion of correct responses would correspond to the bias of the proportion of correct responses to the exercise among all eligible students.

Analyses at this level could conceivably indicate differences in the magnitude of biases among exercises. This information may lead to pinpointing exercises that contribute most heavily to the bias in package by subject matter scores. If some separate exercise bias analyses were done by domain, the primary source or sources of bias might be further pinpointed.

6.2 Design Effects and Power

Two supplementary evaluative measures of analyses are the sampling design effect and the power (or sensitivity) of the tests that are performed. The former indicates the precision of estimates from the study's sampling design relative to a design with the same sample but where simple random sampling is performed. The ratio of the variance from the present design to the simple random sampling design is known as the design effect. The latter measure is defined as the probability that a significant bias is detected by a test, given that the results are of some arbitrary degree of significance.

Since the computation of all analyses provide for computing approximate variances of all estimates according to the present design, design effects are computed by obtaining approximate variances of the biases and rel-biases according to the assumption of simple random sampling. The estimates of biases and rel-biases are represented as combinations of various products, sums, and ratios of random variables which implies that, even under the simplifying sample design assumptions, the variances must be approximated. As one possible set of approximations for the above, if x and y are estimators and $E(x) = X$ and $E(y) = Y$, then

$$Var(x + y) = Var(x) + Var(y) + 2Cov(x,y)$$

$$Var(xy) \doteq X^2 Var(y) + Y^2 Var(x) + 2YXCov(x,y)$$

$$Var(y/x) \doteq \frac{1}{X^2}\left[Var(y) + \frac{Y^2}{X^2} Var(x) - 2\,\frac{Y}{X}\,Cov(x,y)\right] ;$$

Since X and Y will not be known, x and y, respectively, can be used in the above formulations.

The power of hypothesis tests of bias, subject to certain simplifying distributional assumptions and levels of significance, can be illustrated by compiling several tables where precision measures vary among the tables and the magnitude of the bias, whose significance is to be detected by the test, is varied within the tables. The power or sensitivity is ascertained by observing the estimated bias magnitude (assuming it was the true bias being tested as equal to zero) and the variance of the estimate.

6.3 Imputation Procedures

Imputation procedures present a means by which biases attributable to nonresponse can be reduced somewhat, but never eliminated completely. If prior findings indicate a small overall bias, imputation procedures may sufficiently reduce the bias to levels where major analytic studies of nonresponse bias would no longer be necessary. Assuming this scenario to be a realistic one, an investigation of various useful imputation procedures may be rewarding.

Several approaches are possible. If it were found that regular assessment nonrespondents are comparable, by some definition, to regular assessment respondents from the dropout or early graduate supplementary frames, then any nonresponse bias could be reduced by weighting the supplemental frame respondents more heavily to compensate for the data lost to nonresponse. For this reason, a comparison of supplementary frame and No-Show package data has been done. Several comparison combinations of supplemental frames with In-School or Out-of School No-Shows are possible and have been tabulated (see appendix D). The outcome of these and other similar analyses may yield a model by which accurate estimates can be made from regular assessment response data alone, thus obviating the need for nonresponse followup. Another weight adjustment technique is the so-called hot-deck procedure which was used for the 1960 Census. In this method, cross-classification cells are formed by one or more relevant variables available from both respondents and nonrespondents. For example, some of the auxiliary data variables could be used for this purpose. Each cell is supplied with a single response from among the respondents. As the file of data is read, new responses for the cells are supplied as they are encountered. These responses remain until another response from the same cell appears. When a nonresponse is encountered it is identified with a cell. The weight of the nonrespondent is then attached to the current respondent occupying the cell. When several relevant variables are available for both respondents and nonrespondents, the Automatic Interaction Detector subroutine can be used to specify which cross-classifications of these variables should be used to define hot-deck cells or weighting classes for the substitution method mentioned below [9]. Finally, a weight adjustment procedure involving the use of multiple regression may be used. With this technique, the binary response variable is regressed on some set of relevant variables which are available for respondents and nonrespondents. The nonresponse adjustment for each respondent is the inverse of the fitted response variable. The usefulness of the regression approach is contingent upon a high degree of association between the auxiliary and response rates. This approach is similar in nature to the so-called Politz scheme in which the adjustment is the reciprocal of a respondent's probability of being found.

Imputation by substitution of additional selections from the population (i.e., alternates) is currently used for individual nonresponse in NAEP. A selection of additional students is made within schools. These students are then selected to replace students originally selected but who fail to show up. In addition to failing to yield unbiased estimates, it should be noted that adjustment techniques tend to reduce the precision of estimates, assuming all else equal.

Several of these imputation techniques could be tested and compared. The criterion of comparison would be the mean squared error of these estimates. This criterion is computed as the variance of the estimate plus its bias squared. The variance can be found directly and the bias can be estimated using "true" parameter values estimable from regular assessment and No-Show sample data. Data manipulation techniques like the imputation procedures mentioned above are alternatives to more costly followup surveys aimed at No-Shows. A cost model can be developed for a followup sample of No-Shows to judge the relative cost and mean squared error efficiency of survey versus data manipulation methods.

6.4 Pilot Study

6.4.1 Experimental Design

The purpose of this pilot study will be to investigate two novel approaches which could be used to increase the proportion of initially selected students who are administered packages. Schools will be categorized as one of two types, according to the number of days required for assessment. One category will involve those schools requiring one day or less and the other will involve those requiring more than one day. One of the approaches that will be tested (i.e., design treatments) dictates that No-Show followup be done and that makeup session(s) be arranged on the same day as the scheduled administration. The other approach dictates that followup be done and that makeup sessions be arranged on a subsequent day. For completeness, these two approaches will be compared with the present approach used by NAEP (i.e., the control). Group sessions with sixteen selections per session will be used to reflect Year 07 sampling procedures.

6.4.2 Statistical Quality

To assess the level of precision that one should expect from the No-Show followup pilot study proposed in section 6.4.1, the design layout in table 6.1 has been considered.

Table 6.1. Layout of No-Show followup pilot study

Blocking factors		Followup procedures					
		Control		Same day		Next day	
Package assignment load	Expected response level	Sample schools	Response rate - %	Sample schools	Response rate - %	Sample schools	Response rate - %
One day assignment	Low	6	68	3	78	3	85
	High	6	78	3	84	3	88
More than one day assignment	Low	6	68	3	78	3	85
	High	6	78	3	84	3	88

Notice that a blocking factor which categorizes schools according to their anticipated level of response has been incorporated into the layout above. Anticipated low response rate schools will consist roughly of Low Metropolitan, Urban Fringe, and Suburban Fringe schools. High response rate schools will be drawn from the High Metropolitan, Medium City, Small City, and Extreme Rural STOC categories.

Implicit in table 6.1 is the assumption that there will be little or no difference in response rates between schools with a one day package assignment and those with a two or more day assignment. The response rate figures assumed for the control groups are typical of those observed for Year 04 17-year-olds in tables C.13 through C.15 of appendix C. The response rates set for the same day and next day followup procedures reflect the level of improvement that we would hope to achieve. The response rate layout in table 6.1 also assumes that the next day procedure will be superior to the same day followup and that the improvement among high response rate schools will not be as great as that among low response schools.

The statistical power of the pilot study design proposed above will depend on the precision of the estimated response rates. Suppose that i = 1, 2 denotes the low and high response rate schools; j = 1, 2, 3 denote respectively the control, same day, and next day followup procedures; h = 1, 2 depicts the one day assignment and two-or-more-day assignment schools; and k = 1, 2,..., s(ijh) indexes the response level-i schools assigned treatment-j. Suppose further that r(ijhk) denotes the response rate among the n(ihk) students selected for assessment in school-ihk and followed up with procedure-j. The response rate for treatment-j among response level-i schools will be estimated by

$$r(ij) = \sum_{h=1}^{2} \sum_{k=1}^{s(ijh)} r(ijhk)/2s(ijh) \qquad (6.4.2.1)$$

If R(ijhk) is the response rate that would be obtained from all the N(ihk) eligible students in school-ihk when treatment-j is used, then the expected value of r(ij) over successive student selections and randomization of treatments is

$$R(ij) = \sum_{h=1}^{2} \sum_{k=1}^{s(ih)} R(ijhk)/2s(ih) = \sum_{h=1}^{2} R(ijh)/2 \qquad (6.4.2.2)$$

where $s(ih) = 12$ for all ih pairs since there are twelve schools in each of our four experimental blocks. The response rate figures in table 6.1 represent hypothesized values of the $R(ijh)$ parameters. Ignoring finite population corrections and assuming that the student sample sizes $n(ihk) \doteq n(h)$ for all 24 schools in package load block-h, one arrives at the variance expression

$$\mathrm{Var}\{r(ij)\} \doteq \frac{\sigma^2_{sch}(ij)}{s(ij)} + \frac{\sigma^2_{std}(ij)}{s(ij)\left\{\sum_{h=1}^{2} n^{-1}(h)/2\right\}^{-1}} \qquad (6.4.2.3)$$

where

$$\sigma^2_{sch}(ij) = \sum_{h=1}^{2} \sum_{k=1}^{s(ih)} [R(ijhk) - R(ijh)]^2/2[s(ih) - 1]$$

$$\sigma^2_{std}(ij) = \sum_{h=1}^{2} \sum_{k=1}^{s(ih)} R(ijhk)\,[1 - R(ijhk)]/2s(ih)$$

and

$$s(ij) = s(ij1) + s(ij2) = 12 \text{ for } j = 1 \text{ and } 6 \text{ for } j = 2 \text{ or } 3.$$

To produce values of these components consistent with the $R(ij)$ values hypothesized in table 6.1 we have used the model

$$\sigma^2_{std}(ij) = .99 \times R(ij)\,[1 - R(ij)]$$

and

$$\sigma^2_{sch}(ij) = .01 \times R(ij)\,[1 - R(ij)] .$$

For one day assignment schools we assume that the average number of packages assigned is 2 leading to $n(1) = 2 \times 16 = 32$. For two-or-more-day assignments the average number of packages was set to 6 resulting in $n(2) - 6 \times 16 = 96$. The harmonic mean of these student sample sizes is

$$\bar{n} = \left[\frac{1}{2}\left\{\frac{1}{32}+\frac{1}{96}\right\}\right]^{-1} = 48.$$

These considerations lead us to

$$\mathrm{Var}\{r(ij)\} = \left\{\frac{R(ij)\ [1 - R(ij)]}{s(ij)}\right\}\left\{.01 + \frac{99}{48}\right\}$$

$$= (.030625)\ R(ij)\ [1 - R(ij)]/S(ij) \qquad (6.4.2.4)$$

Table 6.2 displays R(ij) and associated V(ij) = Var{r(ij)} values along with contract coefficients C(ij) for 5 single degree of freedom orthogonal contrasts among the r(ij).

Table 6.2. Response rates, variances, and contrasts

Response level	Low response schools			High response schools		
Followup method	Control	Same Day	Next Day	Control	Same Day	Next Day
Hypothesized R(ij)	68	78	85	78	84	88
School sample s(ij)	12	6	6	12	6	6
Variances V(ij)	5.55	8.76	6.51	4.38	6.86	5.39
Same vs. Next	0	-1	+1	0	-1	+1
Followup vs. Control	-2	+1	+1	-2	+1	+1
Low vs. High	-1	-1	-1	+1	+1	+1
(S vs. N) x (L vs. H)	0	-1	+1	0	+1	-1
(F-U vs. C) x (L vs. H)	-2	+1	+1	+2	-1	-1

To calculate the power of students-t tests for the five orthogonal contrasts outlined above, we need the noncentrality parameters

$$\phi = \left|\sum_{i=1}^{2}\sum_{j=1}^{3} C(ij)\ R(ij)\right| \Big/ \left[2\sum_{i=1}^{2}\sum_{j=1}^{3} C^2(ij)V(ij)\right]^{1/2}$$

$$= |CR|/\sqrt{2}\,\sigma_{CR} \qquad (6.4.2.5)$$

and the degrees of freedom (df) available for estimating the denominator of (6.4.2.5).

Letting

$$I(ij) = \begin{cases} 1, & \text{when } C(ij) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

we will use

$$\nu = \sum_{i=1}^{2} \sum_{j=1}^{3} I(ij)[S(ij) - 1] \qquad (6.4.2.6)$$

These considerations lead to the power calculations (two-sided, size $\alpha = .05$) summarized in table 6.3.

Table 6.3. Power of two-sided t-tests

Source	$\lvert CR \rvert$	$2\sigma_{CR}$	ϕ	ν	α	β
Same vs. Next	11	7.48	1.48	20	.05	.50
Followup vs. Control	43	11.60	3.71	42	.01	.99
Low vs. High	19	8.65	2.19	42	.05	.82
(S vs. N) x (L vs. H)	3	7.42	0.41	20	.05	<.10
(F-U vs. C) x (L vs. H)	11	11.60	0.95	42	.05	$\doteq$.30

Table 6.3 demonstrates that the power for our primary followup versus control contrast should be very good. The Low response versus High response contrast has adequate power, the Same versus Next contrast is weak, and the interaction contrasts have very little power.